

法政大学学術機関リポジトリ  
HOSEI UNIVERSITY REPOSITORY

## 「心理学研究法総論」をどう教えるか(1)

著者	吉村 浩一, 関口 洋美, 野川 中
出版者	法政大学文学部
雑誌名	法政大学文学部紀要
巻	78
ページ	165-183
発行年	2019-03-18
URL	<a href="http://doi.org/10.15002/00021791">http://doi.org/10.15002/00021791</a>

# 「心理学研究法総論」をどう教えるか（Ⅰ）

吉村 浩一・関口 洋美・野川 中

キーワード：認定心理士（心理調査）、心理調査概論、心理学研究法、実験法、調査法

## 1 序論

### 1.1 はじめに

心理学の大きな特徴に、「自分の生み出したデータを使って論を組み立てる」ことがある。最近では「メタ分析」と言って、自分自身では新たにデータは生み出さず、公表済みの他の研究者のデータを比較可能な状態に整理した上で筆者の論点を展開する研究スタイルもあるが、そこでもデータの再分析などのために研究法的技能が不可欠である。要は、学部での学習段階から、心理学では、実験・調査・観察法など何らかのデータ収集法とデータ分析法を駆使し、「新たなデータを産出すること」が求められるのである。

「自分で新たにデータを生み出す」作業には、2つのフェイズがある。データを「集める」フェイズと「分析する」フェイズである。学部や大学院で心理学を学ぶ際には両者は異なる科目として授業展開されることが多いが、両者は密接に関連するため、本論ではデータ分析の諸方法をその分析法と関連深いデータ収集法に取り込んで論じていきたい。大まかに言うと、実験法と調査法には量的データに関わる統計的分析法が、観察法と面接法では質的データを整理する分析法に関わる。検査法では、量的・質的の両データ分析法が関わってくる。

心理学研究法を取り巻くこうした事情を背景に、なぜ新たな研究法科目の検討が必要かを説明

することから始めよう。

### 1.2 日本心理学会が認定する「認定心理士（心理調査）」の必須科目

昨今は心理学でも資格問題が賑やかで、よく知られているものに「臨床心理士」がある。これは、日本臨床心理士資格認定協会による協会認定資格であるが、病院や学校を始め、さまざまな臨床場面で実効性をもつ資格である。また、2017年度には、国家資格として新たに「公認心理師」が制定されたが、これも臨床系科目に重点を置いた資格で、全国の数ある心理学科の中には受験に必要な科目群を設定できない心理学科も多い。

長らく日本心理学会が学会認定していた「認定心理士（心理調査）」が加わった。「心理調査に関連する専門科目を履修した認定心理士」（日本心理学会ホームページ）に与えられる資格である。注意すべきことは、ここで言う「心理調査」とは、いわゆる心理学が用いてきたさまざまな研究法の中の調査法だけを指すのではなく、実験法や検査法、観察法、面接法を含めた心理学研究法の総称という点である。この紛らわしさは、のちに述べる開講科目名にも混乱を来たしかねないことを、あらかじめ指摘しておきたい。

法政大学文学部心理学科においても、この「認定心理士（心理調査）」の資格認定を受けるための科目群の申請を、2016年度の制度開始から行ってきた。この資格認定を受けることは、「心理学

科」と名乗る学科であれば、特段、新たな科目を開講するまでもなく、現行カリキュラムでほぼ達成されている。ただし、1科目だけ、新設を要する科目がある。それは、「心理調査」に関わる「心理調査概論」という基本科目である。全国に数ある心理学科の中には、実験法・調査法・検査法・観察法・面接法・統計法など、研究法別の授業科目は設定しているが、研究法全体の概論あるいは総論を独立した科目として設定していないところが少なくない。法政大学文学部心理学科も、そうした総論的な研究法科目は設けてこなかった。この科目の履修は2020年度までに卒業する学生には必須ではなく、個別研究法科目を複数履修することで代替可能であるが、その猶予期間ののち、2021年度以降に卒業する学生に対しては、「心理調査」資格の基本科目として、履修が義務づけられることになる。

科目名の例として「心理調査概論」が示されているが、上述のように、「調査法」とは、心理学では実験法や検査法と並ぶ研究法の中の1つにすぎないため、研究法全体の総称としては不適切である。したがって、この趣旨の科目を設置するにあたり、科目名として「調査」ではなく「心理学研究法」とするのが適切である。

### 1.3 法政大学文学心理学科の心理研究法関連科目群

法政大学文学部心理学科では、心理学研究法に関する科目をかなり充実させている。それらは厳格な意味での必須科目ではないが、「標準的に修得すべき科目」としてほとんどすべての学生が受講している。まず、1年次の春・秋学期には心理学の学習法全般に関して「基礎ゼミⅠ・Ⅱ」、実験法に関して「心理学基礎実験Ⅰ・Ⅱ」、そして統計法に関して「心理統計法Ⅰ・Ⅱ」と、6科目12単位分が開講されている。2年次には春学期に「演習Ⅰ」（学び方全般）、秋学期に「演習Ⅱ」（実験法）、それに春・秋学期に「統計処理技法Ⅰ・Ⅱ」と「心理学測定法Ⅰ・Ⅱ」（調査法）の合計6科目12単位分が開講されている。さらに3年

次には、春・秋学期に「心理検査法Ⅰ・Ⅱ」（検査法）の2科目4単位分が開講されている。研究法に関して3年間でこれだけの科目群を設けているにもかかわらず、「認定心理士（心理調査）」の資格申請に必要な研究法の総論（概論）的科目がない。

求められている科目を設置するにあたり、各論に進む前の入門とすべきか、それとも心理学で用いられている諸方法を一通り習得した後でそれらを整理するための総論とすべきかの方針を決めなければならない。各論に進む前の入門なら、新たな科目を1年次に配当すべきである。それに対し、諸方法を比較し、それぞれの方法の特徴や問題点を浮かび上がらせる内容にするなら、3、4年次での開講が適切である。どちらの趣旨で開講するかは、カリキュラム全体を見渡し、各学年に配当されている研究法関連科目の単位数を考慮して決めるのが現実的である。法政大学文学部心理学科の場合、1、2年次に研究法関連科目をこれ以上増やすと、比重が重くなりすぎる。加えて、1、2年次は（一般教育科目の履修に重点が置かれるため）心理学専門科目の受講科目数に上限が設けられており、1、2年次での開講はいよいよ難しい。したがって、さまざまな研究法を一通り修得した3、4年次での設置が適切である。そこでの開講は、卒業研究などで自らが用いるべき研究法の選択を主体的に行う力を身につけさせる上でも有益である。

第1著者（吉村）は、これまで実験法を中心にさまざまな研究法を駆使してデータ収集を行ってきた。とは言え、学部学生が卒業研究などで用いるすべての方法について論じることはいできない。たとえば、最近の学生がLINEなどのSNSを用いて安易にあるいはチェーンメールのように調査を行っている様子にかねがね疑問を抱いている。インターネットの発達により、心理学に限らずさまざまな領域でネット調査が広まりつつあるが、実施に当たっては満たすべき条件や守るべきルールがあるはずである。第1著者はこの問題に的確にコメントできないので、ネット調査の特徴とそ

れらを適切に行うための条件について、第2著者(関口)の協力を求めた(4.6)。加えて、第2著者とは、調査法の章の執筆にあたり議論を重ね、有益な情報提供を受けた。また、心理学での実験法と統計法を生かした学生の有望な就職先として、医療統計分野がある。その領域での仕事に携わっている第3著者(野川)には、そうした専門性を生かした職務内容を解説してもらった(3.10)。本稿では、これら2名の協力を得て、心理学で用いられているさまざまな研究法の特徴と問題点を浮かび上がらせていくが、前半となる本論文(I)では、5つの主なデータ収集法のうち実験法と調査法を扱い、残る検査法、面接法、観察法については、続編の(II)で取り上げる予定である。

## 2. 心理学で用いるさまざまなデータ収集法

現在の心理学で用いられるさまざまなデータ収集法は、どれも心理学固有の方法ではない。たとえば実験法は、物理学や工学のように物質を対象とする実験とは一線を画せるにしても、医学や生理学など生命を扱う分野での実験法と共通する部分が多い。また調査法は、アンケート調査を主なデータ収集手段とする社会学などと共通する。観察法は、動物学や植物学をはじめ自然科学での重要なデータ収集法である。このように、心理学で用いているデータ収集法は、他の学問とさまざまに関わっている。その上で、心理学で用いられるデータ収集法の主なものを列挙すると、

- ・実験法
- ・調査法
- ・検査法
- ・面接法(インタビュー法を含む)
- ・観察法

がある。これら5つの分類は残念ながら、MECE(ミーシー)とは言えない。Mutually Exclusive Collectively Exhaustive (MECE) とは、その分類がすべてを覆い尽くしており(Collectively Exhaustive)、かつお互いに重複せず背反

していること(Mutually Exclusive)を言う。このことは分類の理想だが、心理学でのデータ収集法の分類は、そのように明快にはいかない。たとえば、次のような重複がある。

- a. 実験法に検査法が混入することがある
- b. 調査法を面接法形式で行うことがある
- c. 検査法を面接法形式で行うことがある

aの例を示そう。実験する際の独立変数の水準設定を、検査法で得たデータに基づいて行うことが珍しくない。たとえば、向性検査の点数に基づいて内向性群と外向性群に分け、それら2群を「向性」要因(独立変数)の2水準として実験を行うなどである。具体例として、向性(独立変数)の違いが実験的に設定した場面での発話量(従属変数)に違いを生むかを検討する実験を考えよう。これを〈例a〉として、まずは実験法から始めたい。

## 3. 実験法

### 3.1 準実験という考え方

前節最後の〈例a〉のように独立変数を設定する実験は、心理学ではむしろ多数派だが、このことは心理学での実験に重要な制約を与える。そもそも実験とは、条件統制のもとに遂行され、独立変数と従属変数のあいだに因果関係を見いだそうとするものである。それには、出発点として複数の等質な群を用意することから始めなければならない。そして、用意した等質な2群のどちらか一方(実験群)にだけ、ある操作を加え、操作を加えなかった群(統制群)と同じ課題(従属変数となる)を課し、従属変数の値(成績)に(有意な)違いが生じるかどうかを調べる。あるいは、2つの等質な群それぞれに異なる操作Aと操作Bを加え、それらを実験群A・実験群Bとして、やはり両群に同じ課題を課して成績を比較する。これが実験の基本スタイルである。要するに、2つの群の違いは、ある操作を加えたか加えなかったか(ないしは加えた操作AとBの違い)だけで、それ以外は両群等質でなければならない。そ

の前提があるからこそ、従属変数に生じた差を独立変数の操作の違いによるとして、因果関係を主張できるのである。

ところが、〈例 a〉では、等質な2群を作ることなく、向性検査の成績の違いに基づいて2群が設けられた。一見すると、2群の違いは向性の違いだけと思えるかもしれないが、2群はほかの重要な事柄に関しても違っている可能性がある。こうした事態で行う実験を、「真性の実験」と区別して「準実験 (quasi-experiment)」(南風原, 2001; 坂元, 2004 など) と呼ぶのである。

### 3.2 実験法の基本枠組み：独立変数と従属変数そして剰余変数

実験とは、「独立変数」を操作し、その水準の違いにより「従属変数」に違いがあるかどうかを調べるものである。しかし、実験の理解は、これら2つの変数だけでは足りず、第3の変数である「剰余変数」について理解しておくことが不可欠である。これら三者の関係について、高野 (2004) は次のように簡潔にまとめている。

- ・独立変数：原因であると推定される変数（実験者が操作する変数）
- ・従属変数：結果であると推定される変数（実験者が測定する変数）
- ・剰余変数：独立変数以外で従属変数に影響すると推定される変数（実験者が統制する変数）

もっとも単純なのは、独立変数と従属変数がともに1つしかない実験である。ただしその場合でも、「剰余変数」は、数も内容も特定できない場合が多い。この厄介な剰余変数を、なんとか統制しなければならないのである。

準実験とは、言ってみれば、この剰余変数が統制されていない状況で行う実験である。〈例 a〉では、独立変数は向性の違いだったが、設定された2つの群は向性以外はすべて等質とは言い切れない。外向性群は内向性群より男子の割合が高いかもしれないし、大都市出身者の割合が高いかもしれない。2つの群のこうした違いが、従属変数

として測定される発話量に影響するかもしれない。準実験は、このような可能性を含んだまま行う実験であるため、真性の実験のもつ「従属変数に認められた違いは独立変数の違いにより生じた」との因果性について断定ができないのである。

物質科学における実験は、「真性の実験」が原則であるが、生命や心を扱う科学においてはそうした状況確保が困難な場合が少なくない。最近では研究倫理上の配慮を優先させる必要も加わり、「真正な実験」を行うことはますます難しくなっている。心を扱う心理学は、むしろ「準実験」が基本だとして、実験法を考えるべきかもしれない。

### 3.3 「真正な実験」を目指して

独立変数の操作を加える前に等質な2群が用意できれば、「準実験」ではなく「真正な実験」が行える。それを実現するもっとも代表的な方法が「無作為化 (ランダム化) 法」である。たとえば、40名の参加者を20名ずつの2群に分ける場合、2群への振り分けそのものを無作為に行う (random assignment)。こうすれば、結果としてあらゆる面で等質な2群になると期待できる。

しかしながら、この方法には条件がある。この方法で等質な2群になると見込めるのは、「大数の法則」に適用する場合のみである。すなわち、参加者数がかなり多い場合にしか、「無作為化 (ランダム化) 法」によって等質な群分けができるとは期待できない。はたして、総数40人を20人ずつの2群に分けることは大数の法則に適用状況と言えるのだろうか。そう考えると、心理学実験でよく見られる総数20名を10名ずつの2群に分ける操作で等質な2群ができるとはとても期待できない。

各群10名程度の限られた参加者しか確保できない場合には、独立変数による操作を加える前に、最低限、操作前における従属変数の値に両群間で大きな差がない (有意差のない範囲に収まっている) ことを確認しておくべきである。

両群間の等質性をさらに踏み込んで一致させる方法に、「被験者マッチング法」がある。それは、従属変数に影響を与える疑いのある剰余変数が1つか2つに特定できる場合に有効な方法である。たとえば、ある実験での従属変数が、課題処理に要する反応時間だとしよう。そして、従属変数の成績には、とにかく刺激を見たらすぐにボタン押しする「単純反応時間」の個人差が影響する可能性があり、それが考えうる主な剰余変数だとしよう。一方の群に単純反応時間の速い人が偏っていれば、効果を誤って評価しかねない。そこで、単純反応時間の個人差という剰余変数（交絡要因）の影響を取り除くため、参加者 20 人の単純反応時間をあらかじめ測定し、1 位から 20 位までの順位づけを行っておく。そして、1 位の人を A 群、2 位の人を B 群に振り分け、続けて 3 位を B 群、4 位を A 群という要領で、10 人ずつの 2 群を作る。そうすれば、単純反応時間の成績に関する限り、両群の平均値に大差のない 2 群ができる。さらに、1 位の人と 2 位の人をペアと見なし、以下対応する順位同士をペアリングしていけば、10 名全体の平均値の等質性だけでなく、両群間で比較する相手（ペア）を特定できることになる。似た値の人同士を合わせる（マッチングすること）から、この方法を「被験者マッチング法」と呼ぶ。この方法を用いれば、群間の差の検定に際しては、ペア同士を「関連する（対応ある）標本」と見なして検定することになる（たとえば、山内, 2009）。また、分散分析を行うのであれば、被験者内要因として扱える。こうすることで、「真正な実験」に近づくとともに、「検定力」を高めることにもなる。

とは言え、〈例 a〉のように、等質性を確保した上で群分けすることができない準実験が、心理学実験では実に多い。「準実験」でのデータ評価に際しては、上述の限界を踏まえ、飛躍した結論は慎まなければならない。

### 3.4 複雑化する実験法

実験のロジックは、本来は単純である。あるこ

とがらの効果（影響）の有無を判定することを目的に、独立変数としてある操作を加える条件（実験条件）とそれを加えない条件（統制条件）を設け、評価の対象となる指標（従属変数）を比較する。その際、評価を公正に行うため、実験群と統制群は、評価したいことがら以外、等質なのが「真正な実験」である。しかし、上で見たように、心理学では等質な群作りが難しい場合が多く、準実験も許容せざるを得ない。たとえそうであっても、実験群と統制群が示す従属変数の値を比較し、両者間に統計的有意差があるかどうかを示すことが実験法の本質である。

「実験群と統制群の比較」から、「2 つの実験群間の比較」へと発展した。〈例 a〉にあった、内向群と外向群はどちらかが統制群ではなく、2 つは水準の異なる実験群である。次に、「異なる水準」が、2 つから 3 つ以上へと拡張した。たとえば、「内向群」「中間群」「外向群」の 3 水準設定である。試験勉強を 1 時間した人たちと 2 時間した人たち、さらに 3 時間した人たちとのあいだでテスト成績（従属変数）を比較する場合などもある。勉強時間という 1 つの要因（独立変数）に関して、3 つ以上の水準を設定する方向への拡大である。

上記の 1 つの変数内で水準数を増やす方向とは別に、1 つの実験内で扱う要因（独立変数）の数自体を増やす方向への拡張も行われた。〈例 a〉で言えば、たとえば向性という要因以外に、性差を第 2 要因として同じ実験内で 2 つ以上の要因の効果を同時に効率よく検討する。1 つの実験で 3 つ以上の水準を比較したり、2 つ以上の要因を組み込んで効率よく検討するため、「実験計画法」と呼ばれるシステマティックな方法が開発された。これは現在の心理学ではあたり前のように用いられているが、実験心理学が生まれたヴェントの頃（19 世紀後半）にはなかった方法である。心理学での利用が広まったのは、第二次世界大戦以降であった。実験計画法は、心理学だけで用いられているわけでない。それはむしろ当然で、「実験計画法」の開発は心理学以外の分野で行われ

た。最近の心理学研究では、後で述べる実験計画法の誤用・濫用が目につく。それに歯止めをかける意味でも、実験計画法がどのように生まれ使用され始めたかを見ておきたい。

### 3.5 実験計画法の生い立ちと適用範囲

実験計画法で得られた実験データは、分散分析、すなわちF検定で統計処理される。実験計画法は、このFの語源となったフィッシャー(R.A. Fisher)が圃場実験でのデータを系統的に評価するために考案したものである。1935年に出版された初版の『実験計画法』は版を重ね、第8版(Fisher, 1966)は1971年に日本語にも翻訳された。圃場実験とは、品種改良や育成法の効果などを評価する農事試験のことである。試験を行うために圃場を小区画のブロックに分け(心理学では1人の実験参加者に相当する)、要因や処理の効果を統計的に検定することになる。これは、今日の心理学ですっかり定着している実験スタイルである。

圃場実験にあたっては、日当たりや土質、水はけなどの条件がそろったブロックを確保するのは難しい。比較的条件のそろった面積を使って、小分割できる区画(ブロック)の数には限りがある。加えて、作量などの評価は年に1度しか行えない(したがって、要因1つずつの実験を順番に行うと何年もかかる)。このことを心理学に置き換えると、少数の実験参加者だけで、一度に複数の要因を効率よく配置して実験を行わなければならないことになる。実験計画法はそのために開発されたのである。限られたブロック数(心理学では実験参加者数)での実験は、「大数の法則」に頼れない。そのためFisher(1966)は、ベアリング(上述のマッチング法)やラテン方格法など、少数のブロックを公平に割り当てる配置を行った。

ベアリング(対にすること)については、すでに3.3の「被験者マッチング法」で紹介したので、ここではラテン方格について説明する。

Fisher(1966)の邦訳書(フィッシャー、

1971)に、次のような記述がある。

農事試験に用いる土地をまとまった形のブロックに分けて、その各ブロックの中で、比較しようとするすべての実験処理を均等に代表させるようにすれば、限られた地域に対して、一定量の努力と管理上の配慮とを費やして得られる実験上の比較の精度は、非常に向上することがわかった。(p. 57)

この記述の意味を、1要因多水準実験を例に説明しよう。たとえば、6水準の処理(AからF)を施すとする。実験に使用できる土地全体を6×6(縦横同数)の小区画に分割し、6つの処理(AからF)を公平かつ効率よく配置するための方法がラテン方格である。

各処理は同じ行と同じ列に一度だけ現れる。たとえば、Aは6行のそれぞれに1度ずつ、かつ6列のそれぞれにも1度ずつ配置されている。次の配置は、この条件を満たしている。

A	B	C	D	E	F
B	C	D	E	F	A
C	D	E	F	A	B
D	E	F	A	B	C
E	F	A	B	C	D
F	A	B	C	D	E

これは、AからFまでのアルファベットを、行が進むごとに1つずつずらせたものである。確かに、AからFの各水準は、各行各列に1度ずつ配されているが、この配置に対してフィッシャー(1971)は、「地味の尾根または帯が行や列を斜めに横切って延びていれば、ある処理は他の処理よりも系統的に有利になるかも知れない」(P. 62)と不備を指摘する。上記の配置は、右上から左下に向かう斜めに同じ条件Fが配置されている。地味の問題は圃場実験での問題だが、これを心理学実験に置き換えると、試行順序の配慮への必要性となる。心理学の場合、各行の並びは、1人の

実験参加者への6つの処理の試行順序となる。たとえば、水準Aは他の水準に比べ、次に行う試行に強い「キャリーオーバー効果（持ち越し効果）」をもっているとしよう。上の配置では、Aの次にBがくることが多く、Bは他の条件に比べ不当に強いキャリーオーバー効果を被ることになる。これから察せられるように、各処理をあらゆる点で公平に配置することは思いのほか難しい。残念ながら、フィッシャー（1971）は完全無作為化とラテン方格の違いを明確に説明していないため、心理学における適切なラテン方格の作り方については、他の研究（たとえば、Bradley, 1958）に委ねたい（<http://rintintin.colorado.edu/~chathach/balancedlatinsquares.html> 参照）。

ここでは実験計画法における配置の公平さに焦点を当てたが、より重要なことは、限られたブロック（心理学の場合は少ない実験参加者）でいかに公正な評価を行う配置を組めるかである。3.4の終わりに予告しておいたが、のちの調査法において、おびただしい数の参加者を投入して行う質問紙調査で、さも実験計画法に基づいて行ったかのような統計処理を行っている研究が最近の卒業論文等で目立つ。ある質問への回答を独立変数操作であるかのように水準化し実験計画法に基づく分散分析をすることは、実験計画法の濫用・誤用と言うべきである。この点については、調査法の章でも改めて取り上げたい。

### 3.6 心理学で用いられる4つの尺度

実験法に限らず、心理学で扱うデータは、通常は言語などの反応の場合が多く、質的データとしての性質が強い。たとえば、実験参加者にある高さの音を聞かせ、「今の音の高さは何Hzでしたか」と定量的回答を求めても、数値での回答は難しい。絶対音感をもつ人なら、「今の音はC#（277 Hz）」などと答えられるかもしれないが、それは例外で、通常、量的反応は期待できない。せいぜい、2つの音を聞き比べて、どちらの方が高いかを答えられるくらいである。

こうした人間の反応（言葉による質的表出）を相手にしている心理学では、物理尺度より数値性の低い反応も測定対象に組み込んでいかなければならない。この弱点を考えると、20世紀半ばにStevens（1946）が区別した4つの尺度は、心理学データの範囲を格段に広げることになった。現在でも用いられている彼の「名義尺度」「順序尺度」「距離尺度」「比例尺度」の4つの尺度は、心理学では安定して利用されているように思われる。

ところが、最近の心理統計の教科書を見ると、この4分類の数値性に関する見解に矛盾点が見受けられる。4つの尺度を、数値性の低い「質的変数」と数値性の高い「量的変数」に分ける際、「順序尺度」をどちらに含めるかについて見解の相違がある。一部の教科書では「名義尺度」と「順序尺度」を「質的変数」としているのに対し（たとえば、山内, 2009; 豊川・柳井, 1982 など）、他の教科書では「順序尺度」を「量的変数」側に含めている（山田・村井, 2004; 吉田, 1998 など）。こうした混乱は、筆者が心理学を学び始めた頃（1970年代）にはなかったと記憶している。確認のため、1960年代から定評ある統計書として用いられてきた2冊の教科書（岩原, 1965と肥田野・瀬谷・大川, 1961）を見ると、4つの尺度の説明に際し、質的・量的変数の区分はなされていなかった。

4つの尺度を質的・量的変数に分ける線引きにこだわることには理由がある。のちの4.3での評定尺度の説明のところで改めて取りあげることになるが、順序尺度である評定値データを用いて因子分析を行うことが許されるかどうかの判断に関わってくるからである。

### 3.7 精神物理学的測定法は古典的方法か？

前項で記したように、人間の言語反応は直接的には質的反応であることが多い。そのため、人が行いやすい反応から量的データを得る工夫が必要になる。そうした反応法の考案は、ヴェントによる実験心理学研究室の誕生（1879年）以前に始まっ



ており、1860年代のウェーバーとフェヒナーによる精神物理学的測定法にまで遡ることができる。彼らは、何種類かの反応方法を考案したが、ここでは「人間が行いやすい反応」に焦点を当て説明していきたい。

精神物理学では、感覚器官に与えられるさまざまな物理刺激（面の明るさや音の大きさ、温度など）の物理量を横軸に、その物理量に対し人がどう感知するかを心理量を縦軸にとり、両者の関係を量的に関数表示することが目指された。とは言え、上で強調したように、人には感知したことを量的に表出する力が乏しい。そこで、精神物理学では、「調整法」「極限法」「恒常法」など、人が表出できる反応法を利用した。坂田（1991）の解説を参考に、これら3方法について理解しよう。

・「調整法」：2つの刺激の一方を一定にし（標準刺激）、もう一方の刺激（比較刺激）を標準刺激と同じと感じられるところに向かい自由に調整させる。通常は参加者自身が刺激の大きさを直接変化させるが、場合によっては参加者の指示に従って実験者が変化させることもある。調整の際には、刺激を変化させすぎたり逆に調整量が小さすぎたりといった誤差（調整誤差）が生じる。そこで、比較刺激の変化の方向を、標準刺激より明らかに大きいところから始めて小さく調整していく試行（下降系列）と、明らかに小さいところから始めて大きく調整していく試行（上昇系列）の両方を課す。これにより、調整誤差を小さくするとともに、「同じ」と判断する刺激の範囲の上限と下限を知ることができる。「主観的等価点」の測定に適している。短時間で多くのデータを集められるというメリットがある一方で、参加者に測定の仕組みがすべて知られてしまう（全知的手続き）ため、故意に反応をゆがめてしまわれる可能性がある。

・「極限法」：増加（上昇系列）または減少（下降系列）させる比較刺激の値を実験者があらかじめ決めておき、一定の方向へ変化する刺激に対し、参加者は（標準刺激と比べて）「大きい」

「同じ」「小さい」のいずれかで答える。それぞれの系列の試行は、たとえば下降系列では、「大きい」→「同じ」→「小さい」と変化したところで打ち切られる。「調整法」と比較すると手間がかかる短所があるが、「主観的等価点」だけでなく「弁別閾」の測定にも用いることができる。しかし、刺激の変化方法が参加者に容易に知れてしまうため（半知的手続き）、次の予測がしやすくなる。

・「恒常法」：提示される比較刺激の値は、極限法の場合と同様にあらかじめ決められているが、提示順序がランダムである点が、極限法と異なる。数段階に変化する比較刺激を20～200回程度ずつランダム順で提示し、それぞれに対する反応を求める。そのため、極限法と比べ変化の幅が大きく設けられ、提示した刺激が求める反応変化点と一致するのは稀である。そこでこの方法では、それぞれの比較刺激の値において、ある反応（たとえば「大きい」）の出現率を求め、その反応の出現率が50%となる点を補間し、その値を変化点と推定することになる。「恒常法」は、データ産出に長時間かつ大量の測定値を要するが、全知・半知の手続きとは異なり、参加者の作為が入りにくい長所がある。「主観的等価点」と「閾値」の測定に適している。

これら3つの方法では、ともに変化させない方の刺激値（標準刺激）をさまざまな物理値に設定することで、横軸の物理量の値を設定し、縦軸の心理量とのあいだの関数関係を捉えることが可能になる。

こうした精神物理学的測定法は、現在では「ミュラー・リヤー錯視」や「重さの弁別閾」、「大きさの恒常性」など、学部での基礎実験でしか出会わないかもしれないが、知覚内容を量的評価する手段として、今日でも実用に耐える方法である（古典的方法としての評価が定まっているため、心理学に関する試験問題に出題されやすい）。しかも、実用性を考えると、測定対象は知覚だけに限られず、たとえば「美しさ」や「好ましきさ」

など、感性や感情などの内的心理事象の測定にも利用できる。「古典的方法」ではあるが、決して過去の方法ではない。

精神物理学の測定法は、これら3種類の古典的なもののほかに、その後開発された方法もある。次節では、それらの中から、「マグニチュード評価法」と「一対比較法」を取り上げたい。

### 3.8 その後開発された精神物理学の測定法

マグニチュード評価法は、4つの尺度の創案者 Stevens (1957) により考案されたものである。この方法の実施に際しては、反応の基準値（「モジュラス」と言う）を提示する方法とそれさえ提示せずに行う方法がある。「モジュラス」とは、基準となるある刺激を「この刺激を100（あるいは50）とし、これから示す刺激に数値を与えてください」などと教示して提示した上で、それぞれの刺激に対して数値表現を求める方法である。これまでの説明で「人は感知したことを物理量（数値）のまま表出することが難しい」と繰り返してきた。にもかかわらず、マグニチュード評価法は、単位は物理量ではないものの、数値での直接的表出を求める。Stevens は、こうした方法でも安定した反応を得られることを証明し、新しい測定法として提案したのである。それは物理量との関係において単純な直線的一次関数にはならず、多くの場合、測定対象ごとに異なる指数をもつべき関数となる。

マグニチュード評価法も、上の3方法と同様、感覚・知覚分野での利用から始まったが、その適用範囲は広がり、たとえば、Kerst and Howard (1978) は、この方法を心的イメージの大きさ測定に利用した。それは、知覚体験した大きさと記憶に基づくイメージ上の大きさとを比較する実験であった。実験参加者（アメリカ人）の課題は、アメリカ合衆国の大陸部の48州の面積の大きさを数値で答えることであった。知覚条件群は、合衆国の地図を見ながら数値表現を行った。一方、記憶群では、数分間地図を見たあと、記憶イメージをたよりに大きさを数値で表現した。両群の評

定値は、州の実際の面積（客観的物理量）を横軸とする関数で表され、それぞれベキ指数 $\alpha$ が計算された。両群とも、得られたデータは精神物理学のベキ関数にうまく当てはまったが、ベキ指数の値は異なった。この結果から、Kerst and Howard (1978) は、知覚による大きさ評価と記憶に基づく大きさ評価は本質的に同じベキ関数をとる同型性をもつが、ベキ指数値が異なるため、2つの表象間には何らかの量的差異があると結論した (Spoehr & Lehmkuhle, 1982 より引用)。この研究のように、マグニチュード評価法も、感覚・知覚領域に限らず、イメージ（記憶像）など、より内的な心的機能にも適用可能である。

これまで紹介してきた測定法は、いずれも心理量を物理尺度上に位置づけること（あるいは物理尺度と関連づけること）を目指すものであった。それに対し、これから述べる「一対比較法」は、心理尺度上での位置づけを目指す方法である。くり返し述べてきたように、人には感知した量を物理量のまま表出することが難しい。たとえば、重さが違う3つの錘を渡され、それぞれ何グラムかを安定して答えることはできない。そればかりか、重さがわずかしき違わない場合には、3つを軽いものから順に並べることさえ容易でない。3つの錘をA, B, Cとしよう。ある参加者が $A < B$ ,  $B < C$ と感じたからといって、必ずしも $A < C$ と感じるとは限らない。両者を直接比べたとき、 $C < A$ と感じることも起こり得る（このような三者関係を「一意性」のない「一巡三角形」と呼ぶ）。このように、3つでさえ整合的に並べるのが容易でないのに、より多くのものを整然と順序よく量的評価することはいよいよ難しい。しかし、取りあげた2つだけを比べて、どちらの方が重いか、どちらの方が軽いかなど、指定された属性について「その2つに限定して比較すること」は、それほど難しくない。人のもつこの能力を利用し、たとえば7つある評価対象を2つずつ総当たりで判断させ、それらの結果を総合して7つ全体の順序と距離を心理尺度上に位置づけようとするのが一対

比較法である。

一対比較法は、Thurstone (1927) の「比較判断の法則」を出発点にいくつかのバリエーションが提案されている。1 人の評定者には多くの組み合わせのうち 1 対のみしか評定させない方法もある。これだと、通りがかった人に評定してもらうなど短い時間で済ませたい場合には都合よい。試食などの場合も、1 人の人に全組み合わせを食べてもらおうと、食傷感が生じて適切に評価できないおそれがある。しかし、わざわざ実験室に来てもらう心理学実験の場合には、たった 1 対の評価しかしてもらわないのでは、いかにも効率が悪い。しかも、おびただしい人数の参加者が必要となり、現実的方法とは言えない。各参加者の拘束時間は少々長くなっても、1 人の参加者に全組み合わせの比較を行ってもらい、少数の参加者で実験を完遂できる方が望ましい。

また、比較に際して、程度の差を考慮せず、どちらを選ぶかだけを尋ねる方法（サーストンの方法）と、「非常に」「かなり」「やや」など程度の違いまで段階設定して答えてもらう方法（シェッフェの方法）がある。心理学実験では、効率よい（量的）データ収集ができるという観点から、シェッフェの一対比較法が有望である。この方法については、「芳賀の変法」や「中屋の変法」などいくつか変法が考案されており（佐藤, 1985 参照）、心理学実験に用いるのにいよいよ便利になった。芳賀の変法も中屋の変法も、各参加者にすべての組み合わせについて程度の違いまで答えてもらう点は同じだが、対になる 2 つの刺激の順序効果に配慮する必要がある場合（たとえば「A が先で C が後」と「C が先で A が後」の両方を行う必要がある場合）には「芳賀の変法」を、その必要なく各対 1 度だけの評価でよい場合は「中屋の変法」を用いる。一対比較法も、開発当初は重さの比較など感覚・知覚領域で用いられ始めたが、現在では官能検査（たとえば、増山, 1989; 神宮・飯田, 2009）を始め、好悪などさまざまな心理属性の評価に適用されている。

1912 年にウェルトハイマーから始まったゲ

シュタルト心理学を、知覚領域の学派だと思い込んでいる人が多い。しかし、その後ウェルトハイマー自身が行った創造的思考の研究や、同世代のケーラーのチンパンジーの知恵試験のように思考研究に拡大された。さらに、第二次世界大戦終了後には、ハイダーやレヴィンらが社会心理学を構築する枠組みにゲシュタルト心理学を据えた。こうした拡張性は、本稿で扱っている方法論についても当てはまり、感覚・知覚領域から始まった各種の精神物理学的測定法は、測定対象を感性や感情など内的な心理機能へと広げていった。

### 3.9 遂行課題と反応指標

あえて批判を込めて言うと、心理学実験で得られるデータは、言語による主観的応答に偏りすぎている。ボタン押しや 7 段階評定なども、広い意味では言語による反応と言える。「はい」の代わりにボタン押しをする、程度を数字で回答する、これらも本質的には言語反応である。言語による反応は、本人が意識・認識できることにしか用いられない。加えて、意識内容を正直に答えるとも限らない。意識にのぼらないことの検出や欺瞞的反応も起こりうることを考えると、百数十年の歴史をもつ実験心理学が言語反応以外の反応として開発してきたさまざまな指標も活用すべきである。

まず、心理学が開発してきた遂行課題として、鏡映描写やベグボードなどの運動課題がある。所要時間や正確さなどの成績が反応指標となる。所要時間の測定だけなら、ゲームなどもっと楽しい課題もあるが、心理学が開拓してきた標準的遂行課題には、次のような特長がある。まず第 1 に、細かい単位の物理量での評価が可能である。これは、成績を客観的に精度よく捉えるために必要である。次に、作業に取り組む時点で、経験や知識などによる個人差の影響を最小限に抑えられる点がある。ボードの穴にベグを刺したり、鏡に映った映像を見ながら細い通路からはみ出さないように鉛筆を進めていくなど、ほとんどの人にとって初めて経験する作業である。そうした作業を用い

ることで、学習開始時点でのスタートラインの個人差を小さく抑えられる。こうした遂行課題は、器用さや学習の速さなど、その課題自体を評価対象にするのみならず、たとえば暗算などの第1課題に対する第2課題として課すことで、第2課題により第1課題が妨害を被る程度を量的に測定する副次課題としての利用もある。

運動成績以外には、次のような指標もある。刺激画面内のどこを見ているか（どこに注意を向けているか）を捉えるため、眼球運動を測定する。また、意識レベルでの欺瞞を見抜くため、さまざまな生理的指標を利用する。ウソ発見器（ポリグラフ測定）と呼ばれるものである。意識的にウソをついたり緊張が高まったりすると、心拍数が増加し、微妙な発汗により皮膚電位抵抗が減少する。また、環境の快適性を客観的に評価するために、非接触方式のサーモグラフが用いられる。快適さの程度を言葉で答えてもらうのではなく、「快適なら発汗量が適度である」「快適なら身体各部の皮膚温が一定の範囲に収まる」「快適なら心拍数が一定の範囲に収まる」など、生理的指標を使って評価することで、本人の意識レベルでは捉えられない（加えて欺瞞のない）心的・身体的状態の把握が可能になる。このような評価法は、ヒューマン・インターフェイスなど人間工学の分野でも活用されている（黒須，1994）。

そして最近では何より、心のはたらきを脳活動から捉えようとする取り組みが広がっている。大脳皮質部位による機能分担と相互連絡、そして統合過程を明らかにするため、古くは脳波や誘発電位が中心的に用いられていたが、近年は脳波を含め、fMRIやNIRSなど脳活動を画像化する精度とモデル化が向上している。脳活動を捉える各種の方法には、時間分解能や空間分解能、それに侵襲性などの点で長短があり、目的に応じた使い分けが行われている。

### 3.10 実験心理学や心理統計の知識と技能を生かして：医療統計業務

実験法の章を終えるにあたり、実験法やその分

野で用いられる統計を学んだことを生かせる職種の具体例として、医療統計分野を紹介する。医療業界の新薬開発や治験関連業務に心理学の知識を直接用いることはないが、医療統計分野の仕事には心理学を学んだことで培われた研究方法論・統計解析・論文作成の能力を生かすことができる。それは「データサイエンティスト」としての職業であり、その仕事は現在の大学・専門学校では育成体制がほとんど整っていないため、人材がいない状況にある。データサイエンティストとは、データを正しく理解してそれを的確に読み解ける人材であり、さまざまな分野の人たちが多少のスキルはもっているが、心理学を学んだ人たちは、科学的研究方法の素養があり、不確かな心理現象を正しく科学的に理解するための方法論、データの扱い方、読み方を教育されているため、データサイエンティストとしての基礎力を備えている。現在、データサイエンスの専門教育体制を整える動きが進められてはいるが、まだしばらくは心理学分野のアドバンテージは続く。心理学を専攻する学生は、研究方法論を学んだことが武器となる就職先として、新薬開発の分野を選択肢に入れてよいのではないだろうか。

医療統計、特に治験関連業務ではSAS社が開発している統計解析ソフトウェアSASを使用することが「暗黙の標準（デファクトスタンダード）」となっている。統計解析を実施する際には、一から計算式を作るのではなく、流通しているソフトウェアを活用することが一般的である。データを提示するにあたっては、使うソフトウェアの計算が正確である保証をとる必要があるが、SASを用いる限り、使用者が正確性を証明する必要がない。RやPythonなどを用いた場合には、どのライブラリやパッケージを用いたのか、その中身の保証などを使用者が担保しなければならない。心理学の統計解析ではSPSSやExcelを使用することが多いが、SASがこれらよりも優れている点として、データの加工を柔軟にできること、大規模なデータ（例えば1ファイル数GB）でも高速に動作すること、そして作業がすべてプログラ

ミングによって実行されることがあげられる。プログラミングに基づくデータ加工と解析結果を出力することは、それまでの手順をすべて追跡できることを意味しており、改ざんの抑止として重要な意味がある。他のソフトウェアでもこれらの作業を実施することはできるが、SASはいずれの点でも高いレベルで提供されており、医療統計分野で求められる使い勝手を満たしている。

治験をはじめとする臨床研究では、「例数設計（サンプルサイズ設計）」を行うことが求められる。これは、何例以上あれば想定した差を統計的に検出することができるかを算出するもので、事前に作成する実施計画書に記載することになっている。例数設計を行い、必要例数が算出されたら、そこに研究中の脱落見込み例（1割など）を加えて、研究に組み入れる目標症例数を決定することになる。例数設計は、治療効果を見るために最低限の例数で実施するために行うもので、これには経済的側面と倫理的側面からの理由がある。例数が多いほど確実に統計的有意差を検出できるようになるが、参加する患者が少ないほど研究費用を削減することができる。こうした点を考慮に入れ、最もコストパフォーマンスの高い症例数を把握することができる。また、医療行為にはリスクの発生が伴う。安全性が確認されていない治療において有効性を求めるには、できる限りその対象者を少なくすべきである。

治験で人に対して実施される段階には大きく分けて3段階（第Ⅰ層～第Ⅲ層）設けられている。第Ⅰ層では、少数の健康成人に低用量の薬剤を用いて、体内での薬物動態やリスクを確認して、用量を決定する。第Ⅱ層では、例数設計に基づく例数によって、有効性を確認する。第Ⅲ層では、さらに多くの患者を対象に安全性を確認する。

例数設計を行うためには、有意水準（ $\alpha$ エラー）、検出力（Power： $1-\beta$ ）、効果量、標準偏差（連続量の場合）の情報が必要である。有意水準は、実際には差がないにもかかわらず差がある確率の許容上限であり、医療分野においても一般的には両側5%（片側なら2.5%）とされる。検

出力は、実際には差があるにもかかわらず差がないと判断してしまう確率である $\beta$ エラーを最大値1から引いた値で表される。検出力は80%～90%程度に設定されることが多い。効果量と標準偏差は、先行研究や予備試験などから想定することになる。

例数設計では上記のような情報収集と整理が必要となるため、必然的に研究を行う意味や価値を確認することとなる。臨床研究では、検証する治療の効果と標準治療（あるいはプラセボ）の効果との差が「臨床的に意味をもつ」必要がある。臨床的に意味のある差は、臨床医や治療ガイドラインなどによって経験的に規定されていくものだが、研究で想定される効果の差が、臨床的に意味がないほど小さいのであれば、価値のない研究を実施していることになり、無駄に患者をリスクにさらしていることになる。心理学においては、そもそも探索的な研究が多く、ほんの少しでも差があれば研究の意味はあると考えるが、微妙な差で（かつ分散も小さくない）統計的有意差を検出するためには、サンプルサイズを相当に大きくする必要がある。数人の参加者のデータをプールして、繰り返し数を増やす方法も考えられるが、そうすると、サンプルサイズが大きくなるほど背景因子の偏りが結果へ強く影響するという懸念が生じる。100人の参加に比べ10人の参加では、参加者背景のバランスが1人異なったときの影響は10倍になる。その場合は影響がありそうな背景情報の均一化を図ったり重みづけで結果を調整するなどの方法が考えられるが、そもそも繰り返し数を増やさざるを得ないほどに小さい効果の差を検証することに意味があるのか、もっと差のある現象・条件はないのか、と再検討することが必要である（小さい差でも検証の価値があるのであれば、参加者を増やすべきである）。このように、計画している研究の価値を確認する意味でも、心理学研究においてもサンプルサイズ設計を取り入れることは有用であろう。

本節でみてきたように、医療統計の分野で研究心をもって開発的仕事に取り組むためには、心理

学の実験法や統計法の知識や技能が間違いなく役立つ。意味を考えずに使い方を習得するのではなく、心理学で学ぶ推測統計学の基礎的な考え方を学ぶことの大切さも併せて感じ取ってもらえたいであろう。

## 4. 調査法

第1章で力説したように、本稿で言う「調査法」とは、実験法や観察法と並ぶ心理学研究法の1つであって、「認定心理士（心理調査）」の資格申請で求められる心理学研究法全体を意味するものでない。「社会調査士」の場合、社会学で用いられるデータ収集法の中心が調査法なので、「調査士」と命名することは適切だろうが、心理学の場合は事情が異なり、調査法は数ある研究法の一部に過ぎない。その点を断った上で、心理学で言うところの調査法について、重要な論点のいくつかを示していきたい。

### 4.1 実態・意識調査から独立・従属変数としての調査データへ

一言で言うと、調査対象者にまつわる事実を答えてもらうのが「実態調査」、意見や態度を答えてもらうのが「意識調査」である。こう二大別すると、心理学では質・量ともに「意識調査」に重点が置かれる。「実態調査」に関することがらは、質問票の「フェイスシート」で答えてもらう程度にとどまり、意見や態度など心の状態を捉える質問を行うのが心理学での調査である。とはいえ、実態調査も意識調査も、それだけでは心理学研究とはなりにくい。仮説を立て、それを検証する枠組みをもつことが望ましい。

しかしこの枠組みは、調査データの分析を誤った方向へ導く危険性をはらむ。学部生の研究などで、ある項目への回答を独立変数に据え、他の項目への回答を従属変数として分散分析している場面をしばしば目にする。だが、項目間の関係は、「クロス集計」で処理するのが基本である。

そもそも、分散分析とクロス集計では、何が違

うのだろう。分散分析における独立変数とは、実験において「研究者が操作する変数」であり、因果関係を捉えるための仕掛けであった（3.2参照）。しかし、調査でのある項目への回答は、研究者が操作した変数ではない。3.4で予告した「最近の心理学において目立つ実験計画法の誤用・濫用」がここに見られるのである。

なぜ、このような濫用が起こるのだろうか。通常、項目間で行うクロス集計は、両項目とも「はい」「いいえ」などのカテゴリー選択肢で、それぞれに属する度数の分布を問題にする。その際の統計的検定には $\chi^2$ を用い、比率の差を検定する。ところが、最近の心理学調査では、各調査項目に対する回答を「はい」「いいえ」ではなく、1から5までの数値（本来は順序尺度）で答えてもらう方式が広まっている。そして、その値を量的変数として点数化し、分散分析において従属変数が満たすべき量的変数と見なすことになる。3.2で紹介した準実験であるとの認識の弱さも加わり、因果関係にまで踏み込んでしまう。

こうした混同は、多変量解析の世界でも、独立変数・従属変数という用語が当たり前のように用いられていることに後押しされている。独立変数・従属変数という用語が使われると、両変数間の因果性の想定が当然視されてしまう。たとえば、重回帰分析では、説明する側の変数は「説明変数または独立変数」と呼ばれ、説明される側は「目的変数または従属変数」と呼ばれる。そして、前者は後者の原因と位置づけられる。こうしたことが、モデルとして提案される独立変数—従属変数の因果的関係を、無条件に当てはめる姿勢へと導いてしまう。

### 4.2 モデルを必要とする多変量解析

多変量解析との出会いが（探索的）因子分析であるという人は少なくないと思う。特段の仮説（モデル）をもたず、数多くの質問に対する多数人数からの回答を行列配置して因子分析すれば、コンピュータが「勝手に」関連ある質問項目を因子として抽出してくれる。因子分析や主成分分析

は、従属変数のない（独立変数だけの）多変量解析であるため、このような進め方となる。仮説を立てて因果関係を捉えようとする実験計画法と比べれば、いかにも他人任せである。しかし、見方を変えれば、先入観や偏見（仮説）をもたずデータが導き出したことに忠実な理解を促すため、他人任せにもそれなりの価値はある。こうした探索的因子分析では、結果次第でその後の考察の方向性も違ってくる。多変量解析の世界とこうして出会った人にとっては、重回帰分析やパス解析、分散共分散構造分析などに戸惑いを感じる。これらの多変量解析では、質問項目を作る段階からモデルを想定して臨まなければならないからである。

前節で取りあげた独立変数と従属変数という用語を用いる重回帰分析について、神宮・土田(2008)は「モデル論の危険性」と題し、次のように指摘している。

私たちは、原因と結果の関係を明らかにするために、モデルを想定して、実際のとの適合度を考える。この適合度を調べる道具が多変量解析ということになる。しかし、多変量解析そのものもモデルに過ぎない。つまり、統計的な分析は、研究者の設定した相関関係を特定の統計的発想の枠組みで検証したにすぎない。(中略)別の言い方をすれば、これが多変量解析の限界あるいは制約と言える。分析の限界を把握しながら、これらの制約の中で、どれだけ有用なモデル構成をするかが多変量解析を用いる際の重要なポイントとなる。(p.37-38)

無心でデータを投入したら「勝手に」因果関係が現れてくるというわけにはいかない。因果関係を想定したモデルをあらかじめ立てて臨まなければならない。そうでなければ、相関関係から得られたデータの解釈可能性は多様さを極める。あらかじめ立てたモデルが間違っていたとき、どの程度のモデル変更までなら許容してよいか、また使用した質問項目は果たしてモデル構築に適切かの判断など、根幹に関わる位置づけが必要である。

そのためにも、調査実施に先立ち、モデルに対する明確な考え方をもって臨まなければならない。

#### 4.3 無作為抽出（ランダム・サンプリング）の非現実性

測定対象となる標本は母集団からの無作為抽出であることが、推測統計学の基本である。このことは、実験法にも当てはまるが、調査法の場合はより切実である。ただし、完全な意味での無作為抽出（単純無作為抽出）は不可能な場合が多いので、系統抽出法、層化抽出法、多段抽出法など実現性を踏まえた無作為抽出法が考案されている。社会学での調査はこうした抽出法を経て行われているはずである。それに対し心理学では、たとえば社会心理学分野でも、無作為抽出の手続きを踏むことは稀である。

卒論研究などの心理学調査は、大学生を対象に行われることが圧倒的に多い。したがって、その調査の母集団は、広く見積もっても「現在の日本の大学生」である。だが、その見積もりは広すぎで、調査をお願いした大学の「現在の在学生」、さらには心理学科の学生を対象にした調査なら「その中の心理学科生」とするのがやっとである。さらに言えば、それさえも無作為性が確保されておらず、協力してもらった授業を受講する学生に対する全数調査であるのが実情である。

統計的検定は推測統計学に基づいているため、たとえ上記のような問題を抱えていても、測定対象となった人たちを母集団からの標本と見なし、有意差検定や区間推定を行うことになる。漠然とした範囲の母集団を想定し、その中身の厳密性を問わないのが、「現在の心理学のパラダイム」と言わざるを得ない。しかし、標本抽出に関する最低限の見識は示すべきである。投稿された論文の査読審査の際などには、その見識が発揮されていると信じたい。たとえば、社会人の意見・態度を扱う研究で、大学生のみを対象に行った調査や、日本人全体の防災意識を扱う調査を、甚大な災害を受けた地域の人たちだけを対象に行った場合など、サンプルの背後に想定される母集団について

て、査読者は標本抽出上の問題点として指摘するはずである。

#### 4.4 サンプルングは人だけでない

人に関するサンプルングの場合は、母集団からの無作為抽出でないことがわかりやすいが、人以外の場合は見えにくい。実験の例だが、同じ条件でのAさんとBさんの遂行成績を(Aの成績) = (Bの成績)という帰無仮説のもと、有意差検定する場合を考えよう。2人それぞれ、同じ条件で試行を20回ずつ行い、その反復データをサンプルとして検定する。AさんとBさんのそれぞれの20試行を、その条件で行う無限回試行からの無作為抽出標本と見なすのである。この実験のとき、たまたまAさんは体調が悪かったとしよう。得られた20試行のデータは、体調不良時のデータに偏っており、決して無限回試行からの無作為抽出ではない。しかし、こうしたことを問題視することはまずない。当然、誤った結論を導く危険性が高まる。

「人」や「試行」の標本抽出は、よほど明らかな、あるいは重大な欠陥がない限り、現在の心理学では無作為抽出することまで求めない。それに対し、「状況のサンプルング」については、少し慎重に扱うべきである。たとえば、その人の対人関係場面における態度を、5つの具体的場面をサンプルとして、それらの総合得点を用いて評価したいとしよう。「ケンカしている場面」「デートしている場面」など、さまざまな対人関係場面が考えられるが、無限にある場面の中から5つの具体的場面を選んで標本として用いたい。質問作成にあたり、われわれはこうした状況選択にどれほどの配慮を行っているだろうか。少なくとも、あらゆる状況を想定して具体的場面を設定しようとしているのか、それともある範囲の状況に限定して場面設定しようとしているのかくらいは自覚して状況選定を行い、その点を明示して研究を進めるべきである。

#### 4.5 評定尺度を用いるときの注意点

最近の調査では、回答方法を、「はい」「いいえ」「どちらでもない」などの名義尺度による選択肢ではなく、「全くそう思わない」から「非常にそう思う」までを5段階や7段階などに区切り、いずれに該当するかを答えてもらう評定尺度法が一般化している。そして、「全くそう思わない」には「1」を、「非常にそう思う」には5ないし7点を与え、それらの数値を平均するなど、順序尺度のはずの数字を距離尺度として扱うことが広く行われている（こうしたことに伴う問題について、かつては「リッカート尺度」「ガットマン尺度」「サーストン尺度」などを分類し、それぞれの使用要件について議論されていた）。本来的には、順序尺度を間隔尺度として用いることは誤用のはずだが、経験的には、以下の田中（1973）の見解にも示されるように、距離尺度的扱いが許容されている。

これらの数値は序数〔順序尺度〕であるから、加算性はないので、平均尺度値を計算することは無意味であるが、……カテゴリー判断の法則を適用すると、仮定が正しい限り距離尺度に変換することができる。それらの結果をみると、上記の例〔評定尺度〕のようなカテゴリーの場合、カテゴリー幅は経験的にはほぼ等しいことが多く、元々序数尺度ではあるが、距離尺度として用いても大過はないことが多い。（p. 98）  
（〔 〕内は吉村による補足）

回答する人の立場に立てば、5段階や7段階の値に、「非常に」「かなり」「やや」などの副詞を添えてあることは日常感覚に基づいて回答しやすいメリットがある。しかし、尺度構成的には、それらが5点や4点、3点という等間隔の意味をもつ点数に適切に対応するかどうかが問題となる。程度を表す副詞の形容語を不用意に用いないようにするべきである。

こういった問題や上記の田中（1973）の引用文



で取り上げられている問題に対する1つの対処法として、VAS (Visual Analogue Scale) を用いる手法がある。これは、両端にだけカテゴリーラベルを与えた線分を示し、回答者が線分上の判断位置にマークするものである。この方法の利点は、カテゴリーラベルに影響されることなく間隔尺度水準のデータが確保できるところにある。欠点としては、データコーディング時にマークされた点までの線分の長さを1つずつ計測する手間が大きいことがある。この方法の開発自体は古く (Hayes and Patterson, 1921)、今日では心理学に限らず、さまざまな領域で用いられている (山田・江利川, 2014)。

VASにより目盛りの設定をアナログ化する方法、すなわち間隔尺度化する方法以外に、設定した5段階や7段階の値を順序尺度のまま因子分析する方法もある。その代表的なものが「カテゴリーカル因子分析」である (この因子分析は、清水 (2016) が開発したエクセルを用いた統計ソフトHADなどで行うことができる)。ここで言うカテゴリーカル (データ) とは質的データのことで、カテゴリーカル因子分析は順序性が保証されたカテゴリーカルデータ、すなわち順序尺度を用いて行う因子分析である。カテゴリーカルデータをこのように理解すると、3.6での「順次尺度を質的データと量的データのどちらに分類すべきか」との疑問に、自ずと回答が出るように思える。順序尺度は、やはり質的データ (カテゴリーカルデータ) に分類するのが適切であろう。

質問への回答方法として評定尺度を用いる際に注意すべきもう1つの問題に、単極尺度と双極尺度の区別がある。単極尺度とは、「全く……ない」と「非常に……である」を両端に置くことを典型に、1つの事柄の「無」から「最強 (最大)」までを何段階かに分けて答えてもらう評定尺度である。たとえば、「あなたは友達とケンカをすることがありますか」という問いに対し、「全くない」から「非常に頻繁にある」までを数段階に区分する。それに対し、双極尺度とは、奇数個の段階を設け、中央に「どちらでもない」「どちらとも言

えない」など、中性ポイントを置き、正反対の両極間で程度を答えてもらう方法である。たとえば、「Aさんの印象はどうですか」との問いに対し、「非常に暗い」から「どちらとも言えない」 (中性点) を経て「非常に明るい」までを、両方向に同じ段階数を設けて問う。このように2つの尺度構成法を対照させると、単極尺度において、中間に「どちらでもない」「どちらとも言えない」を設定している問い方をしばしば見かけるが、上述の筆者の考え方を踏まえれば、単極尺度において中央に「どちらでもない」を置くことは、単極尺度を双極尺度と混同してしまうことになるため、避けるべきである。

単極・双極の2種類の問い方があるということは、両者を適切に使い分けなければならないことを意味するが、田中 (1996) は、基本的に単極尺度を用いることを奨める。その理由は、双極尺度では、対になる言葉の反対性の保証が必要で、その確保が必ずしも容易でないからである。少なくとも、双極尺度は限られたものにしか用いられない。たとえば、「輝いている」の反対は？」と問われても、一義的な言葉はない。それに対して、単極尺度は原則的に何に対しても設定できる (しかしながら、単極尺度として用いたつもり「まったく親切でない」は日常用語としては「非常に不親切」となり、双極尺度上で「非常に親切」の対極を意味するものと受け取られかねないという問題は残る)。最低限守るべきことは、得点の平均をとるなどの数的処理を行う複数の項目に、単極と双極の回答方法を混在させないことである。できれば、同一質問紙はどちらかの評定尺度で統一することが望ましい。

#### 4.6 インターネット調査を適切に行うために

近年、心理学調査において、従来の質問紙調査のように紙を使用するのではなく、インターネットを利用する方式が現れてきた。中には、手軽にできるからと自分の属するSNSで協力者を募り、サンプリングの偏りや研究倫理の問題を軽視して調査を実施している学生もいる。しかし、他の研

究方法と同様に、一定の条件やルールを理解した上で実施しなければ、本来の目的に沿うデータ収集は望めない。本節ではインターネット調査の特徴をあげ、実施する際の留意点を述べたい。

偏りのないデータを収集するため、これまでは訪問面接や郵送、電話などを用いて調査が行われてきた。しかし、これらの方法では、住民台帳などの大規模データを基にランダムに協力者を選定するため、多くの個人情報を取り扱うことになり、近年ますます実施が難しくなっている。これらの方法は回収率が低く人手もいることなどから、マスコミによる調査や政治に関する意識調査などの一部を除き、用いられなくなってきた。一方で、かつては調査対象者の偏りが大きいことから敬遠されてきたインターネット調査が多くなってきた。スマートフォンでもインターネットを利用できるようになり、総務省情報通信政策研究所（2017）の「情報通信メディアの利用時間と情報行動に関する調査」によると、2016年にはネットユーザーは1億84万人と、人口の83.5%が利用している。このような現状から、今やインターネット調査は対象者の偏りが大きいとの批判が当たらなくなってきた。

では、質問紙調査と比較したときに、インターネット調査にはどのような利点があるのだろうか。以下の点があげられる。

#### ①短期間に多くのデータを得られる

質問紙を配布して回収するという手間がなく、一度Web上にあげてしまえば、あとは協力者が好きな時間に回答してアップできる。したがって、調査者・協力者ともに時間の制約がない。また、質問紙を用いて大規模調査をしようとするれば、さまざまな会場での実施が必要になるため調査実施者を複数用意することが必要になるが、その人手も必要ない。

#### ②一定数の確保が容易

目標とするデータ数が確保できるまでWebにあげておくことで、データ数の制御が容易である。また、調査会社に依頼すれば、協力

者数と項目数とで料金が決められるため、目標とするデータ数を確実に確保できる。

#### ③データの電子化が容易

Web上で回答してもらったデータは、すでにデジタルデータになっているため、手入力の必要がない。それに対し、質問紙調査の場合の多くは手入力の必要があり、大量データになればなるほど、入力ミスの問題も無視できなくなる。

#### ④地域性を解消できる

Webが使えるところであればどの地域でも、調査者から距離が離れた地域でも調査可能である。もちろん、質問紙調査でも遠隔地に質問紙を郵送するなどしてその地域の研究者などに協力してもらって実施する例はあるが、研究者が不在の地域でも調査可能になる。

#### ⑤匿名性の確保

訪問面接調査や郵送調査、電話調査では、名簿などによって調査対象者が特定される。しかし、インターネット調査では、個人が特定される可能性が低い。調査会社に依頼した場合には、個人の特定ができない形式でデータ提供される。また、調査者がWeb上で調査協力者を募った場合も、住所や氏名などの記載を求めないほか、データを暗号化できるサイトを利用して作成すれば、個人の特定はできない。ただし、有志で調査協力者を募った場合にはその限りではない。匿名性の確保は、個人の特定に関する弊害解消だけでなく、協力者の本音を引き出せるかどうかにも関係する。訪問面接調査などでは、調査者が対面した状態で調査を行うため、協力者の中には「あまり否定的な回答をしたら悪いかな」などの気づかいをしてしまうこともある。質問紙調査でも調査者が立ち会う場合などは、このような気づかいが発生する可能性がある。それに対し、インターネット調査では、調査者の顔が見えない分、協力者が遠慮なく本音で回答できる可能性が高まる。

このように、インターネット調査には利点

が多い。一方で、調査対象者の選出については考慮すべき点がある。以下では、インターネット調査の際に陥りやすい問題をあげ、それらを踏まえて実施の際の留意点を述べる。

大隅（2008）によれば、インターネット調査には調査対象者の集め方として、公募型と非公募型があると言う。公募型とは、リサーチ会社などに登録されている人が自主参加的に調査協力することを言う。一方、非公募型は、従来の標本抽出法で調査対象者を選び、調査協力の合意・許諾を得た後、インターネットで調査をお願いするという方法である。前者ではなりすましなどの可能性があるが、後者では本人確認がされているため、その可能性は低い。一方、前者は調査者に個人を特定されることはないが、後者では少なくとも参加の有無は特定される。また、前者では調査対象者を選んだ基準が漠然とするが、後者では明確である。どちらも一長一短だが、重要なことは、研究の目的に合った調査対象者を選出できているかどうかである。

筆者も以前、インターネット調査会社に依頼して調査を行ったことがある。目的は18歳以上の人を対象に幅広い層からデータを得ることであった。その際には、年齢構成や性別、地域などを人口の構成比率に合わせてもらうように依頼した。インターネットユーザーであるという縛りを外すことはできないが、これだけでも偏りの解消につながる。さらに偏りを防ぐのであれば、職業や年収などによってスクリーニングを実施することも有効である。

自分が使用しているSNSで調査協力を呼びかけて実施することをやってはならない。もちろん、研究目的がそのSNS登録者を対象としたものであるのならよい。しかし、対象を幅広く取ろうとしているのであれば、やってはならない。SNSユーザーは増えており、上述した内閣府の調査では、2016年のSNS利用者は全体の71.2%、そのうちLINEの使用割合は67.0%と高い。しかし、自分の属しているSNSへの呼びかけでは、似たもの同士に偏ることになる。その集

団は、個人個人のつながりによって構成されたグループであることを忘れてはならない。

本稿では、心理学が用いている主な5つの研究法のうち、実験法と調査法を扱った。残る3つの検査法、面接法、観察法については、続編となる（Ⅱ）で取り上げる予定である。

### 引用文献

- Bradley, J. V. (1958) Complete counterbalancing of immediate sequential effects in a Latin square design. *Journal of American Statistical Association*, 53, 525-528.
- Fisher, R. A. (1966) *The design of experiments*. 8th edition. London: Oliver & Boyd. (フィッシャー, R. A. 遠藤健児・鍋谷清治 (訳) (1971) 実験計画法 森北出版)
- 南風原朝和 (2001) 準実験と単一事例実験 南風原朝和・市川伸一・下山晴彦 (編) 心理学研究法入門：調査・実験から実践まで 東京大学出版会 pp. 123-152.
- Hayes, M. H. and Patterson, D. G. (1921) Experimental development of the graphic rating method. *Psychological Bulletin*, 18, 98-99.
- 肥田野直・瀬谷政敏・大川信明 (1961) 心理教育統計学 培風館
- 岩原信九郎 (1965) 教育と心理のための推計学 日本文化科学社
- 神宮英夫・飯田文子 (2009) 対比較法 大越ひろ・神宮英夫 (編著) 食の官能評価入門 光生館 pp. 80-87.
- 神宮英夫・土田昌司 (2008) わかる・使える多変量解析 ナカニシヤ出版
- Kerst, S. M. and Howard, J. H. (1978) Memory psychophysics for visual area and length. *Memory & Cognition*, 6, 327-335.
- 黒須正明 (1994) 使い勝手を測る 浅井邦二 (編著) ころの測定法 実務教育出版 pp. 357-376.
- 増山英太郎 (1989) 対比較法 増山英太郎・小林茂雄 (1989) センセーション・パリエーション：官能検査への誘い 垣内出版 pp. 146-180.
- 日本心理学会ホームページ 「認定心理士 (心理調査資格)」 <https://psych.or.jp/qualification/researchqualification/> (2018年7月12日閲覧)
- 大隅 昇 (2008) これからの社会調査——インターネット調査の可能性と課題—— 日本健康教育学会誌, 16, 196-205.
- 坂元 章 (2004) さまざまな実験法：現実の制約の中

- で 高野陽太郎・岡隆 (編) 心理学研究法：心を見つめる科学のまなざし 有斐閣アルマ pp. 120-146.
- 坂田勝亮 (1991) 知覚研究における測定 市川伸一 (編) 心理測定法への招待 サイエンス社 pp. 149-188.
- 佐藤 信 (1985) 統計的官能検査法 日科技連
- 清水裕士 (2016) フリーの統計分析ソフト HAD：機能の紹介と統計学習・教育, 研究実践における利用方法の提案 メディア・情報・コミュニケーション研究, 1, 59-73.
- 総務省情報通信政策研究所 情報通信メディアの利用時間と情報行動に関する調査2017 (最終閲覧日: 2018年6月8日) [www.soumu.go.jp](http://www.soumu.go.jp) > 政策 > 白書 > 29年版.
- Spoehr, K. T. and Lehmkuhle, S. W. (1982) *Visual information processing*. New York: Freeman. ( 苧阪直行他 (訳) (1986) 視覚の情報処理 サイエンス社)
- Stevens, S. S. (1946) On the theory of scales of measurement. *Science*, 103, 677-680.
- Stevens, S. S. (1957) On the psychophysical law. *Psychological Review*, 64, 153-181.
- 高野陽太郎 (2004) 実験と観察 高野陽太郎・岡隆 (編) 心理学研究法：心を見つめる科学のまなざし 有斐閣アルマ pp. 20-32.
- 田中 敏 (1996) 実践心理データ解析—問題の発想・データ処理・論文の作成 新曜社
- 田中良久 (1973) 心理学研究法 16 尺度構成 東京大学出版会
- Thurstone, L. L. (1927) A law of comparative judgment. *Psychological Review*, 34, 273-286.
- 豊川裕之・柳井晴夫 (編著) (1982) 医学・保健学の例題による統計学 現代数学社
- 山田一成・江利川滋 (2014) Web 調査における Visual Analogue Scale の有効性評価 東洋大学社会学部紀要, 52, 57-70.
- 山田剛史・村井潤一郎 (2004) よくわかる心理統計 ミネルヴァ書房
- 山内光哉 (2009) 心理・教育のための統計法 (第3版) サイエンス社
- 吉田寿夫 (1998) 本当にわかりやすいすごく大切なことが書いてある初歩の統計の本 北大路書房